Jamming model for the extremal optimization heuristic

Jamming model for the extremal optimization heuristic

Stefan Boettcher[1] **and Michelangelo Grigni**[2]

[1] Physics Department, Emory University, Atlanta, GA 30322, USA
[2] Department of Mathematics and Computer Sciences, Emory University, Atlanta,
GA 30322, USA

E-mail: sboettc@emory.edu and mic@mathcs.emory.edu

Abstract
Extremal optimization, a recently introduced meta-heuristic for hard
optimization problems, is analysed on a simple model of jamming. The model
is motivated first by the problem of finding lowest energy configurations for
a disordered spin system on a fixed-valence graph. The numerical results for
the spin system exhibit the same phenomenology found in all earlier studies of
extremal optimization, and our analytical results for the model reproduce many
of these features.

PACS numbers: 02.60.Pn, 05.40.−a, 64.60.Cn, 75.10.Nr

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Many situations in physics and beyond require the solution of NP-hard optimization problems,
for which the typical time needed to ascertain the exact solution apparently grows faster than
any power of the system size [1]. Examples in the sciences are the determination of ground
states for disordered magnets [2–5] or of optimal arrangements of atoms in a compound [6, 7]
or a polymer [8]. With the advent of ever faster computers, the exact study of such problems
has become feasible [9, 10]. Yet, with typically exponential complexity of these problems,
many insights regarding those systems still are only accessible via approximate, heuristic
methods [11]. Heuristics trade off the certainty of an exact result against finding optimal or
near-optimal solutions with high probability in polynomial time. Many of these heuristics
have been inspired by physical optimization processes, for instance, simulated annealing [12]
or genetic algorithms [13].

Extremal optimization (EO) was proposed recently [14], and has been used to treat
a variety of combinatorial [15, 16] and physical optimization problems [5]. Comparative

studies with simulated annealing [14, 15, 17] and other Metropolis-based heuristics [18] have established EO as a successful alternative for the study of NP-hard problems, especially near phase transitions [17] that are associated with the most complex instances of such problems [19–24]. Recently, EO has also been successfully applied to Lennard-Jones glasses [6].

In this paper, we elucidate some properties of the EO algorithm with analytical means. We motivate our theoretical model system with a brief study of a disordered spin system on a random graph. The EO applied to finding ground states of this system reveals the same generic properties found for the algorithm previously. From this problem, we can abstract a set of evolution equations which allow a complete analysis of EO as a function of its single parameter, τ, and the system size, n. In particular, an optimal value for τ as a function of n is determined in close analogy with the scaling found numerically in all previous studies [25]. We finish with a discussion of how this model can be used also to investigate alternative versions of EO, or to analytically compare EO with simulated annealing and other local search heuristics.

2. Spin glasses on fixed-valence random graphs

Disordered spin systems on random graphs have been investigated as mean-field models of spin glasses [26] or optimization problems [23, 24, 27, 28], since variables are long-range connected yet have a small number of neighbours. Particularly simple are α-valent random graphs [17, 28, 29]. In these graphs each vertex possesses a fixed number α of bonds to randomly selected other vertices. Specifically, we have used the method described in [30] to generate these graphs which are also referred to as α-regular graphs. (Note that self loops or double connections are not allowed, and disconnected graphs are highly unlikely.) Just as on a lattice, one can assign a spin variable $x_i \in \{-1, +1\}$ to each vertex and couplings $J_{i,j} \in \{-1, +1\}$ to existing bonds between neighbouring vertices i and j. Then the energy of the system is the difference between violated and satisfied bonds,

$$H = - \sum_{\{bonds\}} J_{i,j} x_i x_j. \tag{1}$$

It is more convenient to consider a linearly related quantity, which merely tallies the number of violated bonds per spin in a configuration

$$e = \frac{H}{2n} + \frac{\alpha}{4} \geqslant 0 \tag{2}$$

where we have used the fact that each graph has a total of $\alpha n/2$ bonds.

Clearly, for all $J_{i,j} \equiv 1$ the spin system has two ferromagnetic ground states with $e = 0$ that are easy to find (all $x_i = 1$ or all $x_i = -1$). But for anti-ferromagnetic bonds $J_{i,j} = -1$, the ground state energy depends on the disordered structure of the graph itself. Only if all loops in the graph were of even length (as in a hyper-cubic lattice), there are again simple ground states, each with an alternating spin pattern (Néel state). Instead, in a random graph, the disorder creates loops that have an equal chance to be odd or even length. Thus, on average, half of the loops can have all bonds satisfied, the other half will have at least one bond frustrated. Since the length of loops in random graphs typically diverges with $\log(n)$, each odd loop almost certainly has other odd loops as neighbours to share a violated bond with. In fact, even for a spin glass, $J_{i,j} \in \{-1, +1\}$, the same argument should hold, since only half of the loops will be frustrated and neighbouring frustrated loops can share violated bonds. We find that the average ground state energies found for either bond distribution are identical for $n \to \infty$, in support of the above argument, but the results appear to differ in next-to-leading order corrections.
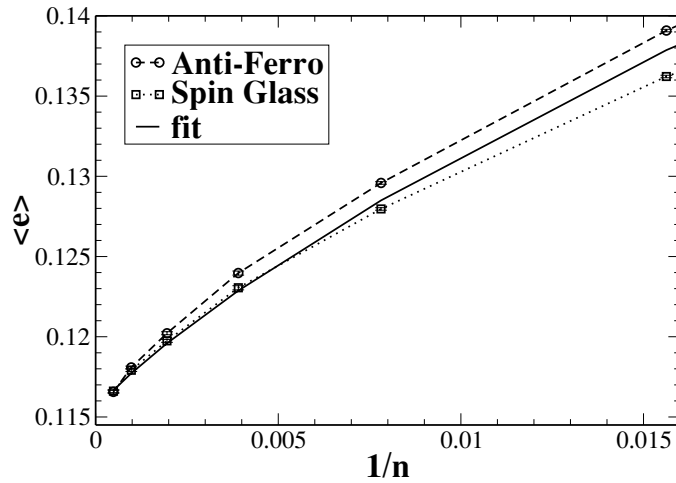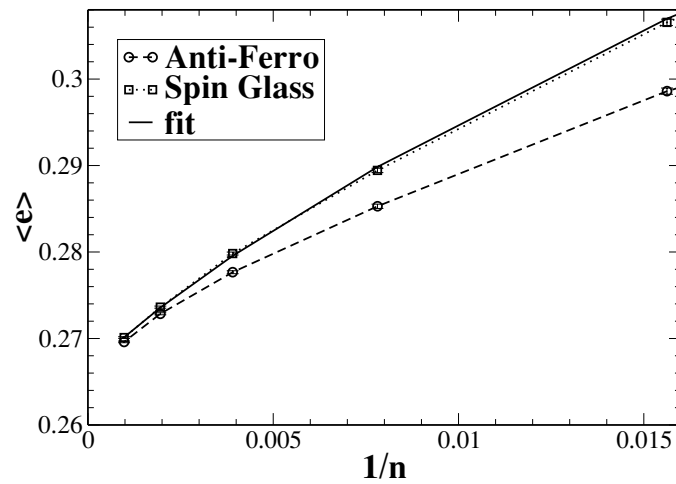
Figure 1. Extrapolation for the number of violated bonds per spin, e, as a function of $1/n$ (a) for trivalent and (b) for tetravalent graphs of size $n = 32, 64, \ldots, 1024$. Circles refer to an anti-ferromagnetic and the squares to a $\pm J$ bond distribution. The error bars for $\langle e \rangle$ are smaller than the symbols. The data for the spin glass are independent of the way the α-valent graph was formed and is best fit (continuous line) by $e_{\alpha=3}(\infty) = 0.1155(5) + 0.35 \ln(n)/n$ and $e_{\alpha=4}(\infty) = 0.266(1) + 0.63 \ln(n)/n$. We found that the data for the anti-ferromagnet for smaller n vary strongly with the way the α-valent graph was formed (here we used the method described in [30]) and are difficult to fit. It is apparent, though, that the difference between the data for the spin glass and the anti-ferromagnet is decreasing for $n \to \infty$.

2.1. τ-EO algorithm for α-valent graphs

To obtain the numerical results in figure 1, we used the following implementation of τ-EO (see also [5]). For a given spin configuration on a graph, assign to each spin x_i a 'fitness'

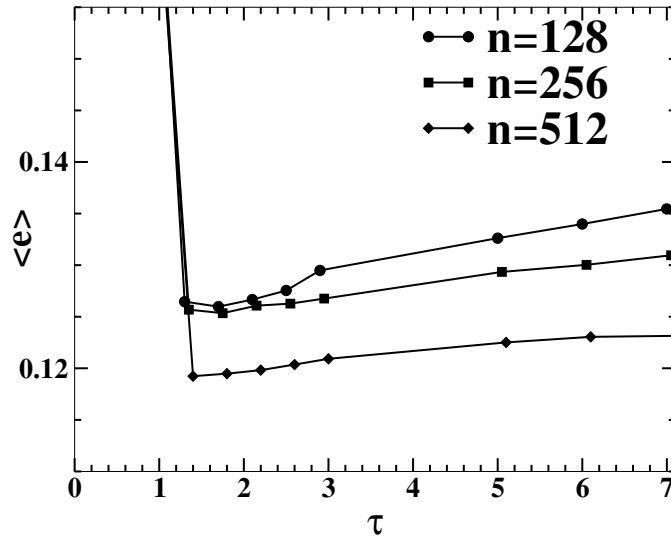$$\lambda_i = -\text{number of violated bonds} = -0, -1, -2, \ldots, -\alpha \tag{3}$$

Figure 2. Plot of the number of violated bonds per spin $\langle e \rangle$ as a function of τ as obtained by τ-EO for a $\pm J$-spin glass on trivalent graphs. Shown are the results for e averaged over 4 runs each on a set of 20 graphs for $n = 128$, 256 and 512. While the results clearly get worse rapidly for $\tau \leqslant 1$, even for $\tau \ll 1$ a decline in the quality can be observed. Despite the slow variation with τ, the value of τ_{opt} where $\langle e \rangle$ is minimized clearly decreases towards $\tau = 1^+$ with increasing n, consistent with equation 6).

so that

$$e = -\frac{1}{2n} \sum_i \lambda_i \qquad (4)$$

is satisfied. Each spin falls into one of only $\alpha + 1$ possible states. Say, currently there are n_α spins with the worst fitness, $\lambda = -\alpha$, $n_{\alpha-1}$ with $\lambda = -(\alpha - 1)$, and so on up to n_0 spins with the best fitness $\lambda = 0$. $\big($Note that $n = \sum_i n_i$.$\big)$ Now draw a 'rank' k according to the distribution

$$P(k) = \frac{\tau - 1}{1 - n^{1-\tau}} k^{-\tau} \qquad (1 \leqslant k \leqslant n). \qquad (5)$$

Determine $0 \leqslant j \leqslant \alpha$ such that $\sum_{i=j+1}^{\alpha} n_i < k \leqslant \sum_{i=j}^{\alpha} n_i$. Finally, select any one of the n_j spins in state j and reverse its spin *unconditionally*. As a result, it and its neighbouring spins change their fitness. After all the effected λ's and n's are re-evaluated, the next spin is chosen for an update.

This EO implementation updates spins with a (τ-dependent) bias against poorly adapted spins on behalf of equation (5). This process is 'extremal' in the sense that it focuses on atypical variables and forms the basis of the EO method. The only adjustable parameter in this algorithm is the power-law exponent τ. For $\tau = 0$, randomly selected spins get forced to update, resulting in merely a random walk through the configuration space. The search is ergodic but yields poor results. For $\tau \to \infty$, only spins in the worst state get updated which quickly traps the update process to a small region of the configuration space which may be far from a near-optimal solution. Ergodicity is broken in the sense that configurations far from the initial conditions are unlikely to be reached within a given runtime. The dependence of performance on τ for this and all previous implementations of τ-EO (for quite different optimization problems [5, 25]) exhibits the features shown in figure 2: The best average

performance in approximating ground state energies at a fixed runtime is obtained for a value of τ_{opt} slightly larger than 1 and $\tau_{\text{opt}} \to 1^+$ for $n \to \infty$. (Note that $\tau = 1$ would be a poor choice! Practical values for, say, an $n = 1000$ random graph typically range from $\tau_{\text{opt}} \approx 1.1$ for spin glasses to $\tau_{\text{opt}} \approx 1.6$ for bipartitioning.) In fact, the (more extensive) numerical data presented in [25] suggested a simple argument that yields

$$\tau_{\text{opt}} \sim 1 + \frac{\ln\left(\frac{a}{\ln n}\right)}{\ln n} \qquad (n \to \infty, \ln(n) \ll a \ll n) \tag{6}$$

where $t_{\text{max}} = an$ was used as the maximum number of updates for a single EO run. This asymptotic behaviour was justified by placing τ_{opt} at the 'edge to ergodicity,' a point between having τ large enough to descend into local minima while having τ just small enough to not get trapped inside the basin of any local minimum. In the following we present a model to make this notion more concrete.

3. Evolution models

We can abstract the random glass problem in section 2 into a simple model which demonstrates previous observations about τ-EO in an analytically tractable way. Consider the spin system on an α-valent graph. Each spin i can be in one of $\alpha + 1$ states λ_i: either no adjacent bond is violated and i is among the n_0 spins or only one bond is violated placing it among the n_1 spins, and so forth up to the n_α spins which have all their adjacent bonds violated. Thus, one can define densities for each of the $\alpha + 1$ states, $\rho_i = n_i/n$ $(i = 0, \ldots, \alpha)$. In general, we can interpret any local search procedure, which only updates a single variable at a time, simply as a set of evolution equations for the $\rho_i(t)$, to wit

$$\dot{\rho}_i = \sum_j T_{i,j} Q_j. \tag{7}$$

Here, Q_j is the probability that a spin in state j gets updated and the matrix $T_{i,j}$ specifies the net transition to state i given that a spin in state j is updated. Note that conservation of probability requires

$$\sum_j Q_j = 1 \tag{8}$$

and conservation of variables requires

$$\sum_i T_{i,j} = 0 \qquad (0 \leqslant j \leqslant \alpha). \tag{9}$$

Both $\mathbf{T}$ and $\mathbf{Q}$ may generally depend on the $\rho_i(t)$ as well as on t explicitly. (For instance, for simulated annealing with a temperature schedule, the Q_i could depend explicitly on t through the changing temperature.)

Another relation is provided by the constraint

$$\sum_i \rho_i(t) = 1 \tag{10}$$

which implies $\sum_i \dot{\rho}_i = 0$. Thus, one of the equations in (7) is always redundant. The cost per variable to be minimized in equation (2) now reads

$$e = \frac{1}{2} \sum_i i \rho_i \geqslant 0 \tag{11}$$

with $e = 0$ being optimal.

The advantage of this notation lies in the fact that the average update *preference,* $\mathbf{Q}$, is separate from the update *process* described by $\mathbf{T}$. For instance, for a random walk (equivalent to τ-EO at $\tau = 0$ or simulated annealing at high temperature) $Q_j(t) \equiv \rho_j(t)$, since the probability that a spin in state i gets chosen for an update is equal to the fraction of those spins, no matter how that update is processed by $\mathbf{T}$. What is typically unknown for a hard problem is the general form of $\mathbf{T}$. But to understand the properties of a heuristic expressed through $\mathbf{Q}$, it may be revealing to 'design' interesting $\mathbf{T}$.

3.1. Annealed approximation to the glass problem

We can construct $\mathbf{T}$ for the glass problem in section 2 on a trivalent graph in an annealed approximation. Since $\mathbf{T}$ in this case is quite messy, and of no great consequence beyond this section, we focus on one of its components, say, $T_{1,2}$. This component represents the net flux in or out of ρ_1, given that a variable in state 2 gets updated. The annealed approximation consists of the unbiased assumption that each of the $\alpha = 3$ neighbouring vertices can be in state i with probability ρ_i independently. Of course, no neighbouring vertex can be in state 0, if the bond to it is violated, or in state α, if the bond to it is good.

For $T_{1,2}$, the vertex chosen for an update has two violated bonds and $\alpha - 2 = 1$ good bonds. First, when that vertex flips, there is a shift of one variable (fraction $1/n$) from ρ_2 to ρ_1. The neighbouring vertex on the other end of each of the violated bonds could be in state 1, 2, or 3 with probability ρ_1, ρ_2, or ρ_3, respectively, and the vertex attached via the good bond could be in state 0, 1, or 2 with probability ρ_0, ρ_1, or ρ_2, respectively. Considering all allowed combinations, we can find the relative (unnormalized) influx into any of the ρ_i as a consequence of updating the vertex at the centre. The sum of the influxes should equal the fraction of moved vertices, α/n, and the relative influxes can be normalized accordingly. Finally, one can identify for each of the combinations where that fraction of moved vertices originated from, which leads to negative outflux to the $T_{i,2}$ (which is obviously required to satisfy equation (9)). The outflux out of state i must be proportional to ρ_i. Thus, we obtain the following three terms contributing to $T_{1,2}$:

$$T_{1,2} = \frac{1}{n} + \frac{\rho_0\rho_1 + 2\rho_1\rho_2 + \rho_0\rho_3 + 2\rho_2{}^2 + 3\rho_0\rho_2}{n(1 - \rho_0)(1 - \rho_3)} - \frac{(3\rho_1 + 3\rho_2 + \rho_3 + 2\rho_0)\rho_1}{n(1 - \rho_0)(1 - \rho_3)} \tag{12}$$

and the construction of the other elements of $\mathbf{T}$ in this annealed approximation proceeds equivalently.

It is not too hard to obtain some steady-state ($\dot{\rho} = 0$) results for equation (7) with this particular $\mathbf{T}$, supplemented by equation (10). One example would be the random walk limit, $Q_i = \rho_i$, equivalent to $\tau = 0$. More revealing for the analysis of EO is the $\tau \to \infty$ limit. In that case, on each update only one among the worst spins gets flipped. From some random initial conditions, EO would empty out state 3 first ($Q_3 = 1$, $Q_2 = Q_1 = Q_0 = 0$), than empty out 2, and so on, until a steady state is reached with the highest non-empty state being ρ_j with some $j > 0$. In this steady state, we can try to determine the $\rho_i(\infty)$ with the ansatz $\bar{Q}_i = c_i$, $\sum_i c_i = 1$, where the average is taken over time. The only consistent balance is obtained with state 3 totally empty, $\rho_3(\infty) = 0$ and $c_3 = 0$, and state 2 almost empty except for a single spin reaching the state sometimes, i.e. $\rho_2(\infty) \approx 0$ and $c_2 > 0$. Hence, $c_0 = 1 - c_1 - c_2$ and $\rho_1 = 1 - \rho_0$, which leads to drastically simplified equations:

$$\begin{aligned}
0 &= c_2(3 + 2\rho_0) + c_1(2 + \rho_0) - (1 + 3\rho_0) \\
0 &= c_2(1 - 4\rho_0) - c_1(1 + 2\rho_0) - (3 - 6\rho_0) \\
0 &= -c_2(3 - 2\rho_0) + c_1\rho_0 + 3(1 - \rho_0)
\end{aligned} \tag{13}$$

all other equations being redundant. The solution is simply

$$\rho_0(\infty) = 1 \qquad \rho_1(\infty) = 0 \qquad c_1 = \frac{1}{2} \qquad c_2 = \frac{1}{2} \tag{14}$$

consistent with numerical simulation for all initial conditions. Thus, in the steady state, almost all variables are in the ground state except for a single vertex that is being bounced between states 1 and 2.

The result that EO converges to the ground state for $\tau = \infty$, while reassuring, is not very helpful to understand either EO or the original problem. The annealed approximation has eliminated everything that made the problem interesting, and EO's convergence for $\tau = \infty$ to a perfectly optimized ground state clearly does not resemble our numerical results from section 2.

3.2. Models with very simple flows

Our naive annealed approximation has eliminated most of the relevant features of the original, hard problem. Not surprisingly, it also fails to predict the existence of a finite value for τ_{opt} (see figure 2); it is easy to convince oneself that $\tau = \infty$ is in fact the best case scenario for τ-EO for all initial conditions and even at finite runtime. Yet, two basic features of the evolution equations remain appealing: (1) The behaviour of a system with a large number of variables can be abstracted into a relatively simple set of equations, describing their dynamics with a small set of unknowns, and (2) the separation of update process, **T**, and update preference, **Q**, lends itself to an analytical comparison between different heuristics. This distinction is possible, of course, only as long as these heuristics can utilize the same single-variable, local search process in **T**. The question is: Can we construct interesting processes **T** in the sense that they capture salient features observed for local search on real, NP-hard problems? We will show that even the most basic versions of **T** provide some insights into the workings of various local search heuristics.

For simplicity, we choose α as small as possible for the three following model situations. Without restriction of generality, in these cases $\alpha = 2$ is sufficient, but more complicated phenomena could be accommodated with more states. First, we consider the most trivial case where a variable when updated merely moves from state i to state $i - 1$ for $i > 0$, or from state 0 to state α (to make every state accessible), $T_{i,j} = \left[-\delta_{i,j} + \delta_{i,(\alpha+j \bmod \alpha+1)}\right]/n$. This process is conveniently depicted as a flowchart in figure 3(a). Clearly, any gradient descent method will be able to reach the ground state $e = 0$ for this process, since there are no barriers. For instance, simulated annealing with zero temperature will reach this state in $O(n)$ trials, and τ-EO for $\tau = \infty$ will reach $e = 0$ in $<n$ steps, when averaged over initial conditions. (Note that in the above notation, $c_i = 1/4$ solves the steady-state equations where $c_0 > 0$ implies $\rho_0(\infty) = 1$, $\rho_{i>0}(\infty) = 0$.) Again, $\tau_{opt} = \infty$ is obvious. In fact, this model can be solved readily for any τ with the methods to be developed below in section 3.3. For the random walk limit, $\tau = 0$, it is $c_i = \rho_i(\infty) = 1/4$ since $Q_i \equiv \rho_i$.

Next, we can reverse the directions of transitions in the previous example to obtain a less trivial case, which now possesses energetic barriers. Here $T_{i,j} = \left[-\delta_{i,j} + \delta_{(\alpha+i \bmod \alpha+1),j}\right]/n$, as depicted in figure 3(b). Remarkably, the previous analysis for τ-EO (at least, for $\tau = 0$ or ∞) does not change. The $e = 0$ steady state is reached again in $<n$ steps for $\tau = \infty$, since EO does not reject uphill moves which are required here to arrive at state 0 through state 2, and $\tau_{opt} = \infty$ again. On the other hand, it is quite clear that simulated annealing will not arrive at $e = 0$ with finite probability in polynomial time, even for a sophisticated temperature schedule. Such energetic barriers are, of course, an inherent feature of many NP-hard problems, which makes this simple model quite revealing.
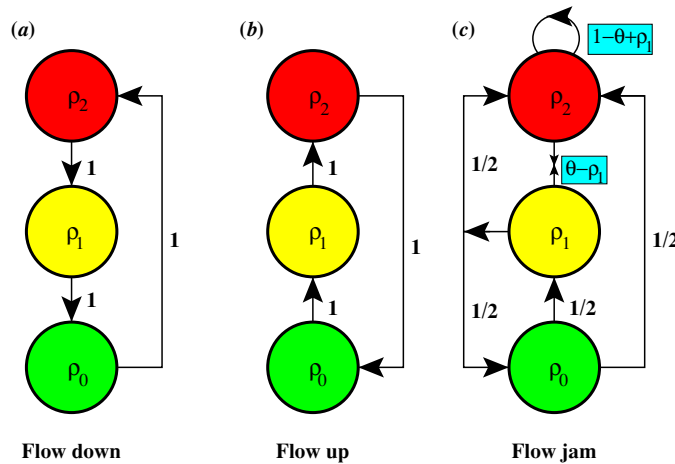
Figure 3. Plot of the flow diagrams for different models discussed in the text. Diagram (*a*) shows a situation in which variables in higher states always evolve towards lower states (except for the lowest state flowing up). In diagram (*b*), variables have to jump to higher energetic states first before they can attain the lowest state. Diagram (*c*) shows the model of a jam, where variables in the highest state can only traverse through the intermediate state to the lowest state, if the intermediate state moves its variables out of the way first to keep its density ρ_1 below the threshold θ. The states have energies that increase from the bottom up, the ρ mark the occupation density of each state and arrows out of a state indicate the rates $nT_{i,j}$ at which a variable flows from state j into another state i, *if* a variable in state j gets updated.

3.3. Model with jammed flow

Naturally, the range of phenomena found in a local search of NP-hard problems is not limited to energetic barriers. After all, so far we have only considered constant entries for $T_{i,j}$. Therefore, in our next model we want to consider the most simple case of **T** depending linearly on the ρ_i. Most of these cases reduce to the phenomena already discussed in the previous examples. An entirely new effect arises in the following case, also depicted in figure 3(*c*):

$$\dot{\rho}_0 = \frac{1}{n}\left[-Q_0 + \frac{1}{2}Q_1\right] \qquad \dot{\rho}_1 = \frac{1}{n}\left[\frac{1}{2}Q_0 - Q_1 + (\theta - \rho_1)Q_2\right]$$

$$\dot{\rho}_2 = \frac{1}{n}\left[\frac{1}{2}Q_0 + \frac{1}{2}Q_1 - (\theta - \rho_1)Q_2\right] \qquad 1 = \rho_0 + \rho_1 + \rho_2. \tag{15}$$

Aside from the dependence of **T** on ρ_1, we have also introduced the threshold parameter θ. In fact, if $\theta \geqslant 1$, the model behaves effectively like the previous models and for $\theta \leqslant 0$ there can be no flow from state 2 to the lower states at all. The interesting regime is the case $0 < \theta < 1$, where further flow from state 2 to state 1 can be blocked for increasing ρ_1, providing a negative feedback to the system. In effect, the model is capable of exhibiting a 'jam' as observed in many models of glassy dynamics [31–33], and which is certainly an aspect of local search processes.

We proceed to calculate the unique fixed point of the system for EO with arbitrary τ. In the general case, the Q depend on the ρ in a more complicated way. As described in the numerical simulation of the glass on a random graph in section 2, with each update a spin is selected based on its rank according to the probability distribution in equation (5). When a rank k ($\leqslant n$) has been chosen, a spin is randomly picked from state α, if $k/n \leqslant \rho_\alpha$, from state

$\alpha - 1$, if $\rho_\alpha < k/n \leqslant \rho_\alpha + \rho_{\alpha-1}$ and so on. We introduce a new, continuous variable $x = k/n$, approximate sums by integrals, and rewrite $P(k)$ in equation (5) as

$$p(x) = \frac{\tau - 1}{n^{\tau-1} - 1} x^{-\tau} \qquad \left(\frac{1}{n} \leqslant x \leqslant 1\right) \tag{16}$$

where the maintenance of the low-x cut-off at $1/n$ will turn out to be crucial. Now, the average likelihood that a spin in a given state is updated is given by

$$Q_\alpha = \int_{1/n}^{\rho_\alpha} p(x) \, dx = \frac{1}{1 - n^{\tau-1}} \left(\rho_\alpha^{1-\tau} - n^{\tau-1}\right)$$

$$Q_{\alpha-1} = \int_{\rho_\alpha}^{\rho_\alpha + \rho_{\alpha-1}} p(x) \, dx = \frac{1}{1 - n^{\tau-1}} \left[(\rho_{\alpha-1} + \rho_\alpha)^{1-\tau} - \rho_\alpha^{1-\tau}\right]$$

$$\ldots \tag{17}$$

$$Q_0 = \int_{1-\rho_0}^{1} p(x) \, dx = \frac{1}{1 - n^{\tau-1}} \left[1 - (1 - \rho_0)^{1-\tau}\right]$$

where in the last line the norm $\sum_i \rho_i = 1$ was used. These values of the Q completely describe the update preferences for τ-EO at arbitrary τ.

Inserting the set of equations (17) for $\alpha = 2$ into the model in equations (15), we obtain

$$\dot{\rho}_0 = \frac{1}{n(1 - n^{\tau-1})} \left[-1 + \frac{3}{2}(1 - \rho_0)^{1-\tau} - \frac{1}{2}\rho_2^{1-\tau}\right]$$

$$\dot{\rho}_1 = \frac{1}{n(1 - n^{\tau-1})} \left[\frac{1}{2} - \frac{3}{2}(1 - \rho_0)^{1-\tau} + \rho_2^{1-\tau} + (\theta - \rho_1)\left(\rho_2^{1-\tau} - n^{\tau-1}\right)\right] \tag{18}$$

$$\dot{\rho}_2 = \frac{1}{n(1 - n^{\tau-1})} \left[\frac{1}{2} - \frac{1}{2}\rho_2^{1-\tau} - (\theta - \rho_1)\left(\rho_2^{1-\tau} - n^{\tau-1}\right)\right]$$

$$1 = \rho_0 + \rho_1 + \rho_2.$$

We abbreviate $A = (1 - \rho_0)^{1-\tau}$ and $B = \rho_2^{1-\tau}$ to obtain for the steady state, $\dot{\rho} = 0$:

$$0 = -1 + \tfrac{3}{2}A - \tfrac{1}{2}B \qquad 0 = \tfrac{1}{2} - \tfrac{3}{2}A + B + (\theta - \rho_1)(B - n^{\tau-1})$$

$$0 = \tfrac{1}{2} - \tfrac{1}{2}B - (\theta - \rho_1)(B - n^{\tau-1}) \qquad \rho_1 = A^{1/(1-\tau)} - B^{1/(1-\tau)}. \tag{19}$$

One of the first three equations is redundant, and we obtain

$$0 = \tfrac{3}{2}(A - 1) + \left[\theta - A^{1/(1-\tau)} + (3A - 2)^{1/(1-\tau)}\right](3A - 2 - n^{\tau-1}) \tag{20}$$

where

$$\rho_0 = 1 - A^{1/(1-\tau)} \qquad \rho_2 = (3A - 2)^{1/(1-\tau)} \qquad \rho_1 = 1 - \rho_0 - \rho_2. \tag{21}$$

The implicit equation (20) has some remarkable properties. It has a single physical solution for the ρ for all $0 \leqslant \tau \leqslant \infty$, $0 < \theta < 1^3$ and all n. In particular, in the thermodynamic limit $n \to \infty$ a critical point at $\tau = 1$ emerges. If $\tau < 1$, the n-dependent term in equation (20) vanishes, allowing A, and hence the ρ, to take on finite values, i.e. $e > 0$. If $\tau > 1$, the n-dependent term diverges, forcing A to diverge in kind, resulting in $\rho_0 \to 1$ and $\rho_i \to 0$ for $i > 0$, i.e. $e \to 0$. This behaviour of $e(\tau)$ for various n is shown in figure 4.

Having a unique fixed point solution seems to be the last word on this problem, with $\tau = \infty$ again being the most favourable value at which the minimal energy $e = 0$ is reached for sure. But it can be shown that the system has an ever harder time to reach that point, requiring typically $t = O(n^\tau)$ update steps for a measurable set of initial conditions. Thus, for

---

[3] Actually, for $\theta \leqslant 0.385$, there is a transition to having three solutions near $\tau = 1$, resulting in a first-order transition for $n \to \infty$. We will only discuss $\theta > 0.385$ here.

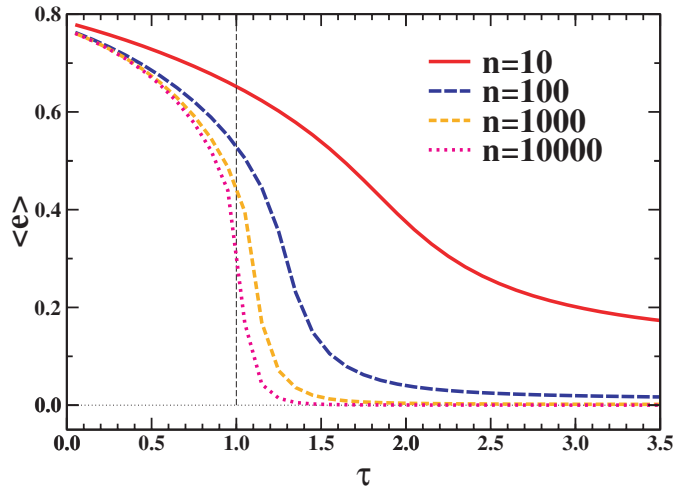**Figure 4.** Plot of $e = \sum_i i\rho_i/2$ as a function of τ resulting from the solution of equations (20) and (21) for $\theta = 1/2$ and various values of n. For $n \to \infty$, a sharp transition emerges at $\tau = 1$, giving optimal results $e \to 0$ for all $\tau > 1$. But for $\tau > 1$ this steady state is reached only for suitable initial conditions, or after sufficient time, see figure 6.

a given finite computational time t_{max} the best results are obtained at some finite value of τ_{opt}. In that, this model provides a new feature—slow variables impeding the dynamics of faster ones [34]—resembling the observed behaviour for EO on real problems, e.g. the effect shown in figure 2. In particular, this model provides an analytically tractable picture for the relation between the value of τ_{opt} and the effective loss of ergodicity in the search conjectured in [14, 25].

The generic evolution of the jamming model for $\tau > 1$ is as follows: if the initial conditions place a fraction $\rho_0 > 1 - \theta$ already into the lowest state, most likely no jam will emerge, since $\rho_1(t) < \theta$ for all times, and the ground state is reached in $<n$ steps. But if initially $\rho_1 + \rho_2 > \theta$ and τ is sufficiently larger than unity, EO will drive the system to a situation where $\rho_1 \approx \theta$ by transferring variables from ρ_2 to ρ_1. Then, further evolution becomes extremely slow, delayed by the τ-dependent, small probability that a variable in state 1 is updated ahead of all variables in state 2. The typical situation in that case is depicted in figure 5.

Clearly, the jam is *not* a steady-state solution of the evolution equations in (18)[4]. But figure 5 suggests the right asymptotic approach to evaluate the intermediate-time behaviour of the jam: consider that we start with initial conditions leading to a jam, $\rho_1(0) + \rho_2(0) > \theta$. We can assume that

$$\rho_1(t) = \theta - \epsilon(t) \tag{22}$$

with $\epsilon \ll 1$ for $t \lesssim t_{jam}$, where t_{jam} is the time at which ρ_2 gets small and the jam disappears. To determine t_{jam}, we apply equation (22) to the evolution equations in (18) to get

$$\dot{\rho_0} \sim \frac{1}{n^\tau}\left[1 - \frac{3}{2}(1 - \rho_0)^{1-\tau} + \frac{1}{2}\rho_2^{1-\tau}\right]$$
$$0 = \tfrac{1}{2} - \tfrac{3}{2}(1 - \rho_0)^{1-\tau} + \rho_2^{1-\tau} - \epsilon n^{\tau-1} \tag{23}$$
$$1 = \rho_0 + \theta + \rho_2.$$

---

[4] It is not even a meta-stable solution since there are no energetic barriers. For instance, simulated annealing at zero temperature would easily find the solution in $t = O(n)$ without experiencing a jam. In reality, a hard problem would most certainly contain combinations of jams, barriers and possibly other features.
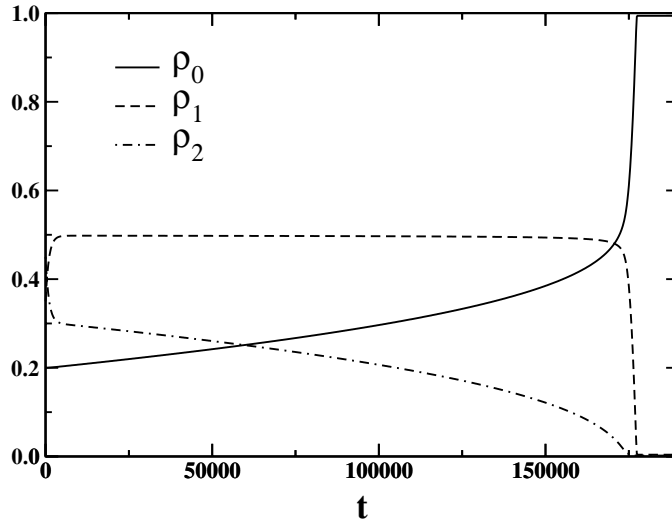
Figure 5. Plot of the typical evolution of the system in equations (18) for some generic initial condition that leads to a jam. Shown are $\rho_0(t)$, $\rho_1(t)$ and $\rho_2(t)$ for $n = 1000$, $\tau = 2$, $\theta = 0.5$ and initial conditions $\rho_0(0) = 0.2$, $\rho_1(0) = 0.35$, and $\rho_2(0) = 0.45$. Since $\rho_1(0) < \theta$, ρ_1 fills up to θ almost instantly with variables from ρ_2 while ρ_0 stays constant. After that, $\rho_1 \approx \theta$ for a very long time ($\gg n$) while variables slowly move down through state 1. Eventually, after $t = O(n^\tau)$, ρ_2 vanishes and EO can empty out ρ_1 directly which leads to the ground state $\rho_0 = 1$ ($e = 0$) almost instantly.

Here, we have already dropped one of the equations (for $\dot{\rho}_2$) which was redundant. Now, we also can disregard the equation containing ϵ, its importance being that it determines the first-order correction, $\epsilon = O(n^{1-\tau})$, *consistently* as a function of the leading order contributions of $\rho_0(t)$ and $\rho_2(t)$. Using the last (norm) equation and its derivative to leading-order, $\dot{\rho}_0 = -\dot{\rho}_2$, we finally obtain an equation solely for $\rho_2(t)$ as

$$-\dot{\rho}_2 \sim \frac{1}{n^\tau} \left[1 - \frac{3}{2}(\theta + \rho_2)^{1-\tau} + \frac{1}{2}\rho_2^{1-\tau} \right] \tag{24}$$

or, using the fact that ρ_2 almost instantly takes on the value of $\rho_1(0) + \rho_2(0) - \theta = 1 - \theta - \rho_0(0)$ (see figure 5)

$$t \sim n^\tau \int_{\rho_2(t)}^{1-\theta-\rho_0(0)} \frac{2\mathrm{d}\xi}{2 - 3(\theta + \xi)^{1-\tau} + \xi^{1-\tau}}. \tag{25}$$

We can estimate the duration of the jam, t_{jam}, by setting $\rho_2(t_{\mathrm{jam}}) \approx 0$, see figure 5

$$t_{\mathrm{jam}} \sim n^\tau f_\tau(1 - \theta - \rho_0(0)) \tag{26}$$

where we defined

$$f_\tau(x) = \int_0^x \frac{2\mathrm{d}\xi}{2 - 3(\theta + \xi)^{1-\tau} + \xi^{1-\tau}} \qquad (x \geqslant 0). \tag{27}$$

Thus, the duration of the jam scales with n^τ times a constant that depends on τ, θ and the initial conditions. As stated before, if the initial conditions keep $\rho_1(0) + \rho_2(0) < \theta$, most likely there will be no jam, reflected in the fact that $f_\tau(x)$ goes to zero for $x \to 0$. The asymptotic scaling in equation (26) conforms well with our numerical simulations: with $f_2(0.3) \approx 0.16$ and $n = 1000$ we obtain $t_{\max} \approx 1.6 \times 10^5$, in good agreement with figure 5.

The long-lived jams that occur for $\tau > 1$ will have a significant effect on the outcome of a local search with EO, which proceeds merely with a finite runtime t_{max}. For instance, for $t_{max} = O(n)$ there are always some initial conditions for which the jam cannot be resolved before t_{max}, resulting in $e > 0$. Thus, the τ-EO implementation faces two conflicting priorities: on the one side, larger τ increases the quality of the steady-state result for e, away from the random-walk-like behaviour at $\tau < 1$, see figure 4. On the other side, $\tau > 1$ increases the chance to get locked into a jam and never to reach that steady state in finite runtime, see figure 5. In between these conflicting interests, we find a preferred value for τ_{opt} that averts both the jam and the random walk, such that $\langle e \rangle$, averaged over initial conditions, is minimized.

Let us assume we fix the runtime to be $t_{max} = an$, where a is a constant with $a \ll n$, so that $n < t_{max} \ll n^\tau$ for $\tau > 1$. If we had chosen $\tau < 1$ for our implementation, there are no jams but we are sure to obtain less than optimal results for $\langle e \rangle$ as in figure 4, so we will assume $\tau > 1$. In this case, we have to distinguish between three possible outcomes to a single run of the EO algorithm, depending on the initial conditions: (1) if $\rho_1(0) + \rho_2(0) < \theta$, the run will most certainly reach the optimal state, $e = 0$, within t_{max} updates, (2) even if $\rho_1(0) + \rho_2(0) > \theta$ but $t_{max} \gtrsim t_{jam}$ from equation (26), $e = 0$ may be reached. Only if (3) $\rho_1(0) + \rho_2(0) > \theta$ *and* $t_{max} \ll t_{jam}$ are satisfied, the search will get stuck in a state of $e > 0$, with a value that depends on the initial conditions. Averaging over all initial conditions, we find

$$\langle e \rangle \approx \frac{1}{\mathcal{N}} \int_0^1 \mathrm{d}\rho_0 \, \mathrm{d}\rho_1 \, \mathrm{d}\rho_2 \, \delta(1 - \rho_0 - \rho_1 - \rho_2)$$

$$\times \frac{1}{2} \left(\sum_{i=0}^2 i \rho_i \right) u(1 - \theta - \rho_0) \, u\left(f_\tau (1 - \theta - \rho_0) n^\tau - t_{max} \right) \qquad (28)$$

where $u(x)$ is the Heaviside step-function and $\delta(x)$ is the Dirac delta-function. The norm is given by

$$\mathcal{N} = \int_0^1 \mathrm{d}\rho_0 \, \mathrm{d}\rho_1 \, \mathrm{d}\rho_2 \, \delta(1 - \rho_0 - \rho_1 - \rho_2) = \frac{1}{2}. \qquad (29)$$

Hence, we obtain

$$\langle e \rangle \approx \frac{3}{2} \int_0^{\max\{0, 1 - \theta - f_\tau^{-1}(t_{max}/n^\tau)\}} \mathrm{d}\rho_0 \, (1 - \rho_0)^2. \qquad (30)$$

The average energy $\langle e \rangle$ in equation (30) will start to rise for increasing τ as soon as the upper integration limit becomes nonzero, or when $t_{max} \approx f_\tau(1 - \theta)n^\tau$. If $t_{max} \ll f_\tau(1 - \theta)n^\tau$, i.e. for $\tau \gg 1$, equation (30) predicts for the average energy $\langle e \rangle = (1 - \theta^3)/2$.

Since $\langle e \rangle$ will reach its minimum value right before the onset of jams causes its rise, we can use this relation to estimate the optimal value of τ. In effect, this justifies the connection between τ_{opt} and the 'edge of ergodicity' noted in [14]. Since the dependence of f_τ on τ is much weaker than the exponential n^τ, we can write

$$\tau_{opt} \sim \frac{\ln(t_{max}/f_\tau(1 - \theta))}{\ln n}$$

$$\sim 1 + \frac{\ln(a/f_\tau(1 - \theta))}{\ln n} \qquad (31)$$

where we have used our choice $t_{max} = an$. In recognition of the fact that $\tau \to 1^+$ for $n \to \infty$, we can simplify the last expression by expanding $f_\tau(x)$ in that limit to get

$$f_\tau(x) \sim \frac{2}{\tau - 1} \int_0^x \frac{\mathrm{d}\xi}{\ln \left[\frac{(\theta + \xi)^3}{\xi} \right]} \qquad (\tau \to 1). \qquad (32)$$
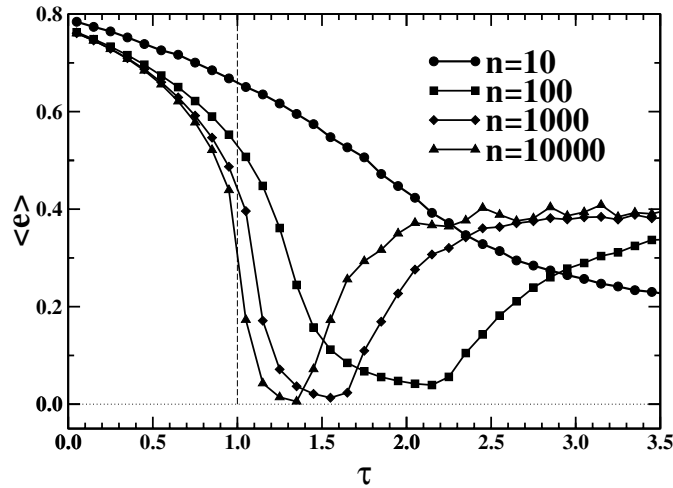
Figure 6. Plot of the energy $\langle e \rangle$ averaged over many τ-EO runs with different initial conditions as a function of τ for $n = 10$, 100, 1000 and 10 000 and $\theta = 1/2$. For small values of τ, $\langle e \rangle$ closely follows the steady-state solutions plotted in figure 4. It reaches a minimum at a value near the prediction for $\tau_{\text{opt}} \approx 3.5$, 2.1, 1.6 and 1.4 and rises sharply beyond that. It reaches an asymptotic value approaching the prediction of $\langle e \rangle \approx 0.44$ for $\tau \to \infty$.

Note that the pole at $\tau = 1$ is a generic consequence of equation (9), independent of the choice of the particular $T_{i,j}$ depicted in figure 3(c). If we insert equation (32) into equation (31), we *exactly* reproduce the n dependence given in [25], see equation (6) above. Numerically, we get at $\theta = 1/2$ for $f_\tau(1/2) \approx (2 \ln 2 - 1)/(\tau - 1)$, and using $a = 100$ and $n = 10$, 100, 1000 and 10 000, equation (31) predicts $\tau_{\text{opt}} \approx 3.5$, 2.1, 1.6 and 1.4.

We can compare this prediction with numerical simulations of τ-EO applied directly to the jamming system described in figure 3(c) (not just the evolution equation in (18) that uses averaged probabilities Q). In figure 6 we show the results for $\langle e \rangle$ as a function of τ for $n = 10$, 100, 1000 and 10 000 at $t_{\text{max}} = 100n$. Initially, for $\tau \lesssim 1$, $\langle e \rangle$ reaches the steady-state result from figure 4 for any initial condition. However, as predicted, $\langle e \rangle$ reaches a minimum at a τ_{opt} beyond which it starts to rapidly deviate from the steady-state solution. This is the 'ergodic edge' beyond which unresolved jams affect the observed value of $\langle e \rangle$. Our prediction for τ_{opt} appears to become increasingly accurate for $n \to \infty$. Furthermore, for $\tau \to \infty$, equation (30) predicts $\langle e \rangle \sim 7/16 \approx 0.44$ for $\theta = 1/2$, in reasonable agreement with the numerical value seen in figure 6.

4. Conclusion

We have presented a simple model to analyse the properties of local search heuristics. This model was applied to extremal optimization and we found conditions under which EO exhibited the same phenomenology on the model as it does on real combinatorial optimization problems as exemplified here by a frustrated spin system on a random graph. The analytical results from the model in equation (31) closely resemble equation (6), the prediction from [14, 25].

Of course, the model is tailored more towards understanding the EO mechanism and does not nearly represent all of the features of a hard optimization problem. (After all, it takes EO only $O(n^\tau)$ updates to find the ground state in the worst case for the model.) Thus, finding a non-trivial value for τ_{opt} in the model merely provides an analogy. For instance, τ_{opt} is

somewhat dependent on the relationship of t_{max} to n. If $t_{max} = O(n^l)$ then $\tau_{opt} \sim l^+$ for $n \to \infty$ according to the model. In this regard the analogy seems to hold in every respect for the graph bipartitioning problem in [14, 25] where $t_{max} \sim n$. But in our numerical simulations for spin glass systems in section 3.2 displayed in figure 2, or in [5], typically $O(n^{3-4})$ updates were required to obtain consistent results for increasing n, yet $\tau_{opt} \to 1^+$ was found irrespective.

We believe that our observation for the behaviour of EO is quite robust under variation of the entries for **T**. More complicated choices for **T** (which may be analytically less tractable) could be made to represent hard problems more closely. In this sense, the separation between **T** and **Q** allows us to study comparisons between different update modes, and even with other local search procedures. As our examples in figures 3(a)–(c) show, simple choices of **T** can lead to interesting scenarios, although there is no real frustration. For instance, one could analyse the properties of EO for different choices of the probability distribution over the ranks in equations (5, 16).

It is more difficult to construct the Q for simulated annealing. Let us assume that we consider a variable in state j for an update. Certainly, Q_j would be proportional to ρ_j, since variables are randomly selected for an update. The Boltzmann factor $e^{-\beta \Delta E_j}$ for the potential update of a variable in j, aside from the inverse temperature $\beta(t)$, only depends on the entries for $T_{i,j}$

$$
\begin{aligned}
\Delta E_j &= n \Delta e_j \\
&= \frac{n}{2} \left[\sum_i i \rho_i(t+1) - \sum_i i \rho_i(t) \right]_j \\
&= \frac{n}{2} \sum_i i T_{i,j}
\end{aligned}
\tag{33}
$$

where the subscript j expresses the assumption that a variable in state j is considered for an update. Hence, we find for the average probability of an update of a variable in state j

$$
Q_j \propto \rho_j \min \left\{ 1, \exp \left[-\beta \sum_i i T_{i,j} \right] \right\}
\tag{34}
$$

which is still short of a proper normalization. Similarly, comparisons with other methods such as threshold annealing [35] can be considered.

Acknowledgments

References

[1] Garey M R and Johnson D S 1979 *Computers and Intractability, A Guide to the Theory of NP-Completeness* (New York: Freeman)
[2] Pal K F 1996 *Physica* A **223** 283
    Pal K F 1996 *Physica* A **233** 60
[3] Hartmann A K 1999 *Phys. Rev.* B **59** 3617
    Hartmann A K 1999 *Phys. Rev.* E **60** 5135
[4] Palassini M and Young A P 2000 *Phys. Rev. Lett.* **85** 3017
[5] Boettcher S and Percus A G 2001 *Phys. Rev. Lett.* **86** 5211
[6] Blair D L and Gould H Private communication

[7] Bhattacharya K K and Sethna J P 1998 *Phys. Rev.* E **57** 2553

[8] Tuzel E and Erzan A 2000 *Phys. Rev.* E **61** R1040

[9] Palmer R G and Adler J 1999 *Int. J. Mod. Phys.* C **10** 667

[10] Desimone C, Diehl M, Jünger M, Mutzel P, Reinelt G and Rinaldi G 1995 *J. Stat. Phys.* **80** 487

[11] Rayward-Smith V J, Osman I H and Reeves C R 1996 *Modern Heuristic Search Methods* (New York: Wiley)

[12] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **220** 671

[13] Holland J 1975 *Adaptation in Natural and Artificial Systems* (Ann Arbor, MI: University of Michigan Press)

[14] Boettcher S and Percus A G 2000 *Artif. Intell.* **119** 275

[15] Boettcher S and Percus A G 1999 *GECCO-99: Proc. Genetic and Evolutionary Computation Conf.* (San Francisco, CA: Morgan Kaufmann) p 825

[16] Boettcher S, Grigni M, Istrate G and Percus A G 2002 *Phase Transitions and Computational Complexity* in preparation

[17] Boettcher S 1999 *J. Math. Phys.* A **32** 5201

[18] Dall J and Sibani P 2001 *Comput. Phys. Commun.* **141** 260

[19] Cheeseman P, Kanefsky B and Taylor W M 1991 *Proc. IJCAI-91* ed J Mylopoulous and R Rediter (San Mateo, CA: Morgan Kaufmann) pp 331–7

[20] See *Frontiers in Problem Solving: Phase Transitions and Complexity* 1996 *Artif. Intell.* **81** Special issue 1–2, ed T Hogg, B A Huberman and C Williams

[21] Monasson R, Zecchina R, Kirkpatrick S, Selman B and Troyansky L 1999 *Nature* **400** 133
Monasson R, Zecchina R, Kirkpatrick S, Selman B and Troyansky L 1999 *Random Struct. Alg.* **15** 414

[22] Cocco S and Monasson R 2001 *Phys. Rev. Lett.* **86** 1654
Cocco S and Monasson R 2001 *Europhys. J.* B **22** 505

[23] Ricci-Tersenghi F, Weigt M and Zecchina R 2001 *Phys. Rev.* E **63** 026702

[24] Franz S, Leone M, Ricci-Tersenghi F and Zecchina R R 2001 *Phys. Rev. Lett.* **87** 7209

[25] Boettcher S and Percus A G 2001 *Phys. Rev.* E **64** 026114

[26] Viana L and Bray A J 1985 *J. Phys.* C: Solid State Phys. **18** 3037

[27] Wong K Y M and Sherrington D 1987 *J. Phys. A: Math. Gen.* **20** L793

[28] Mezard M and Parisi G 1987 *Europhys. Lett.* **3** 1067

[29] Banavar J R, Sherrington D and Sourlas N 1987 *J. Phys. A: Math. Gen.* **20** L1

[30] Bollobas B 1985 *Random Graphs* (London: Academic)

[31] Jaeger H M, Nagel S R and Behringer R P 1996 *Rev. Mod. Phys.* **68** 1259

[32] Ben-Naim E, Knight J B, Nowak E R, Jaeger H M and Nagel S R 1998 *Physica* D **123** 380

[33] Ritort F 1995 *Phys. Rev. Lett.* **75** 1190

[34] Palmer R G, Stein D L, Abrahams E and Anderson P W 1984 *Phys. Rev. Lett.* **53** 958

[35] Franz A, Hoffmann K-H and Salamon P 2001 *Phys. Rev. Lett.* **86** 5219